Allergy

WILEY | EAACI

ORIGINAL ARTICLE OPEN ACCESS

Basic and Translational Allergy Immunology

Single-Cell Omics Analysis of Human Basophils Reveals Two Transcriptionally Distinct Populations

Sofia Papavasileiou[1] | Jiezhen Mo[1] | Daryl Boey[1] | Chenyan Wu[1] | Magnus Tronstad[1] | Lucille Margerie[1] | Remi-André Olsen[2] | Jörg A. Bachmann[2] | Jing Hui Low[3] | Jocelyn Ong[3] | Lars Heede Blom[4] | Anand Kumar Andiappan[3,5] | Gunnar Nilsson[1,6] | Joakim S. Dahlin[1]

[1]Division of Immunology and Respiratory Medicine, Department of Medicine Solna, Karolinska Institutet, and Center for Molecular Medicine, and Karolinska University Hospital, Stockholm, Sweden | [2]Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden | [3]Singapore Immunology Network (SIgN), agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore | [4]Allergy Clinic, Department of Dermatology and Allergy, Herlev and Gentofte University Hospital, Denmark | [5]A*STAR Skin Research Labs (A*SRL), agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore | [6]Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Correspondence: Joakim S. Dahlin (joakim.dahlin@ki.se)

Received: 7 August 2025 | **Revised:** 8 November 2025 | **Accepted:** 9 December 2025

Keywords: basophils | heterogeneity | long-read sequencing | scRNA-seq | single-cell omics

ABSTRACT

Background: Basophils are implicated in various diseases including allergies, but a comprehensive single-cell characterization of human basophils has yet to be performed. Here, we aimed to generate a single-cell omics-based reference resource of circulating human basophils, integrating transcriptomic and large-scale immunoprofiling data. We also sought to investigate basophil heterogeneity at the molecular level.

Methods: Circulating basophils were analyzed using cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq). Both short- and long-read single-cell RNA-sequencing platforms were used to capture the transcriptomic data.

Results: CITE-seq enabled accurate identification and profiling of side scatter^{low} lineage^{-} CCR3^{+} FcεRI^{+} basophils. Short-read single-cell RNA-sequencing data revealed two previously unresolved basophil populations, defined by 66 differentially expressed genes and reproducibly identified across donors. Despite the transcriptional differences, the populations displayed similar immunophenotypes based on more than 100 investigated cell surface markers. Long-read single-cell RNA-sequencing analysis confirmed the existence of the two populations and provided further insights into their gene expression profiles.

Conclusions: We present a multimodal single-cell resource that defines two novel transcriptionally distinct basophil populations. This resource, accessible through a user-friendly web interface, constitutes a cellular and molecular reference map for future studies of basophils in health and disease.

1 | Introduction

Single-cell RNA-sequencing technologies allow for profiling of thousands of individual cells in parallel. Large-scale application of these technologies has allowed consortia, such as the Human Cell Atlas [1], to establish reference collections of transcriptomics data of millions of cells. These collections enable the wider research community to gain cellular and molecular insights into the cell types throughout the body. Virtually all broad classes of immune cells have been characterized at the single-cell transcriptional level, including the allergy-associated eosinophil [2–4] and mast cell [5–7] populations. However, a detailed characterization of basophils has yet to be performed at the single-cell transcriptional level.

Immune cell types are typically composed of phenotypically distinct subsets, yet basophil heterogeneity remains incompletely characterized. Earlier efforts to explore the diversity of basophils have, for example, relied on assessments of physical properties [8] or on surface marker profiling using mass cytometry [9]. However, such approaches depend on predefined conditions and marker panels, inherently limiting the resolution with which basophil heterogeneity can be resolved. By contrast, single-cell omics offers an unbiased, high-resolution characterization of cellular states, enabling systematic identification of novel basophil populations based on molecular signatures.

Here, we perform a comprehensive single-cell omics-based analysis of circulating human basophils. We show that cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) analysis [10] detects basophils among circulating leukocytes. Single-cell RNA-sequencing coupled with large-scale immunoprofiling reveals two transcriptionally distinct basophil states and shows that the identified populations harbor similar immunophenotypes. Long-read-based single-cell RNA-sequencing verifies the transcriptionally distinct populations using an independent sequencing approach and provides additional insights into the basophils' gene expression patterns.

2 | Methods

2.1 | Ethics Statement

The Swedish Ethical Review Authority approved the study, and the subjects provided informed consent before sample collection.

2.2 | Sample Processing and Cell Isolation

The samples were from anonymous subjects that fulfilled the requirements to be blood donors in Sweden, which include criteria of being at least 18 years old, weighing at least 50 kg, and being healthy. The buffy coat fraction of peripheral blood from the donors was processed. Low-density leukocytes were isolated using Ficoll-Paque density gradient centrifugation. Any remaining red blood cells were removed using BD Pharm Lyse (BD Biosciences, Franklin Lakes, New Jersey, USA). Subsequent cell isolation steps were intentionally varied between samples to assess the robustness of the single-cell analysis. The specific approaches are described in separate sections.

2.3 | Flow Cytometry

Cells were incubated with antibodies against CD3 (clone HIT3a), CD19 (HIB19), CD14 (M5E2), CCR3 (5E8), FcεRI (CRA-1), CD117 (A3C6E2). 4′,6′-diamidino-2-phenylindole (DAPI) or 7-aminoactinomycin D (7-AAD) was used to exclude dead cells. The fluorescently labeled antibodies and live/dead markers were purchased from Biolegend (San Diego California, USA), BD Biosciences (Franklin Lakes, New Jersey, USA), and Miltenyi Biotec (Bergisch Gladbach, Germany). The flow cytometry data were analyzed using FlowJo software (Version 10.10.0, BD Life Sciences, Ashland, Oregon, USA). The procedures to analyze the protein expression levels of CXCR4, MS4A3, and galectin-10/CLC are detailed in Supplementary Methods.

2.4 | Generation of the Donor A Dataset Presented in Figure 1

Low-density leukocytes were stained with fluorescently labeled antibodies against CD3 (HIT3a), CD19 (HIB19), CD14 (M5E2), CCR3 (5E8), FcεRI (CRA-1), and the oligonucleotide-labeled antibodies TotalSeq-B0391 anti-human CD45 and TotalSeq-B0090 Mouse IgG1 κ isotype control (Biolegend). The stained cells were split into two cell fractions. One cell fraction was stained with biotinylated β2-microglobulin antibodies (2M2) and the other fraction with unconjugated β2-microglobulin antibodies (2M2) (Biolegend). Each cell fraction was then incubated with TotalSeq-B0953 PE streptavidin (Biolegend). Antibody incubation steps prior to the fluorescence-activated cell sorting were performed in the presence of 2 mM EDTA. We sorted side scatter^{low} (SSC^{low}) PE^{+} lineage^{-} CCR3^{+} FcεRI^{+} basophils from the PE-labeled cell fraction (Figure 1A top) and SSC^{low} PE^{-} leukocytes from the unlabeled cell fraction (Figure 1A bottom) using the FACS Aria Fusion system (BD Biosciences). The two cell fractions were pooled and CITE-seq analysis was performed. The cells were maintained at 4°C or on ice after the isolation of low-density leukocytes.

2.5 | Generation of the Donor B Dataset Presented in Figure 2

Low-density leukocytes were split into two fractions. Each cell fraction was incubated with either biotinylated or unconjugated β2-microglobulin antibody (2M2), followed by the addition of TotalSeq-B0953 PE streptavidin together with the fluorescently labeled antibodies against CD117 (A3C6E2), CD3 (HIT3a), CD19 (HIB19), CD14 (M5E2), CCR3 (5E8), and FcεRI (CRA-1). Antibody incubation steps prior to fluorescence-activated cell sorting were performed in the presence of 2 mM EDTA. Unlabeled basophils (SSC^{low} lineage^{-} CCR3^{+} FcεRI^{+} PE^{-}) and PE^{+} c-Kit^{+} basophils (SSC^{low} lineage^{-} CCR3^{+} FcεRI^{+} PE^{+} c-Kit^{+}) were sorted (Figure S1D) using the FACS Aria Fusion system (BD Biosciences). The sorted cells were pooled (Figure S1E) and stained with oligonucleotide-labeled antibodies that included TotalSeq-B0090 Mouse IgG1 isotype control and TotalSeq-B0391 anti-human CD45 before the CITE-seq analysis (Biolegend). The cells were maintained at 4°C or on ice after the isolation of low-density leukocytes. The PE^{+} c-Kit^{+} basophils did not form a transcriptionally distinct population in the bioinformatics analysis (Figure S1E–G). All basophils (PE^{-} unlabeled, PE^{+} c-Kit^{+} labeled, and ambiguously labeled) were therefore presented together in Figure 2 and referred to as side scatter^{low} lineage^{-} CCR3^{+} FcεRI^{+} basophils.

2.6 | Generation of the Donors C and D Datasets

Low-density leukocytes were incubated with CCR3-APC-Cy7 antibodies (Biolegend) followed by staining with anti-Cy7 magnetic beads (Miltenyi Biotec). Magnetic-activated cell sorting enriched the CCR3^{+} cells. The cells were maintained at 4°C during

the staining and magnetic separation. An additional Ficoll density gradient centrifugation was performed at room temperature to remove dead cells. The CCR3$^+$ cells were then maintained at 4°C or on ice during the remaining steps. We incubated the cells with the TotalSeq-B Human Universal Cocktail v1.0 (Biolegend) according to the manufacturer's recommendations. The CCR3$^+$ cells were further enriched and washed using a magnetic column. Flow cytometry analysis using the BD FACSCanto II system (BD Biosciences) was performed on a fraction of the cells to verify cell purity. The antibody incubation and magnetic separation steps were performed in the presence of 2 mM EDTA. However, the EDTA was removed before the CITE-seq workflow as it interferes with the cDNA synthesis.

2.7 | Short-Read Single-Cell Transcriptomics Data Generation

The isolated cells were processed using the 10X Genomics platform (Pleasanton, California, USA). Specifically, single-cell datasets were generated with the Chromium Next GEM Single Cell 3′ Reagent Kit v3.1 (dual index) with Feature Barcode technology for Cell Surface Protein (10X Genomics). To ensure the capture of the basophils, known to have low RNA contents [11], one additional PCR cycle was introduced in the first cDNA amplification step. The final libraries were sequenced on the Novaseq 6000 platform (Illumina, San Diego, California, USA).

2.8 | Bioinformatic Analysis

Illumina sequencing data were analyzed using Cell Ranger (10X Genomics) and Seurat (version 4.3.0.1) [12]. Cells were assigned based on TotalSeq CD45 sequencing reads, whereas barcoded and non-barcoded cells were identified based on TotalSeq PE sequencing reads (Table 1, Figure S1A,B). Sequencing depth and quality control filters for each dataset are presented in Table 1. For details about data processing, see Supplementary Methods.

2.9 | Long-Read Sequencing

Long-read sequencing was performed using the Oxford Nanopore Technology (ONT). For protocol and analysis details, see Supplementary Methods.

3 | Results

3.1 | CITE-Seq Captures the Transcriptome of Basophils at the Single-Cell Level

Gene expression analysis of primary human basophils has been performed in bulk using microarray- and RNA-sequencing-based methods [11, 13, 14]. However, the capacity of the single-cell RNA-sequencing workflow to detect the human basophil population has yet to be evaluated and strategies to confidently annotate the cell population in datasets are missing. We therefore devised a CITE-seq-based strategy to track primary basophils from peripheral blood throughout the experiment, from the input sample to the resulting dataset. Briefly, we sorted fluorescently and oligonucleotide-barcoded basophils (side scatter^{low} lineage$^-$ CCR3$^+$ FcεRI$^+$ cells [15]) and mixed them with low-density leukocytes (Figure 1A). Flow cytometry analysis showed that the cell suspension used as input for the single-cell RNA-sequencing contained 26% barcoded basophils (Figure 1B). As a comparison, 17% of the cells in the single-cell dataset constituted barcoded cells after quality control and filtering (Figure 1C, Figure S1A–C), revealing successful capture of the basophil population with a small selective loss of these cells.

We visualized the single-cell transcriptomics data using Uniform Manifold Approximation and Projection (UMAP) plots (Figure 1D–F). The absence of a single-cell-based reference to annotate the basophil population based on gene expression profile prompted us to specifically plot the oligonucleotide-barcoded cells (Figure 1D). These cells constituted the flow cytometry-defined basophils, allowing for accurate

TABLE 1 | Total number of sequencing reads per sample and quality control filtering cut-offs.

Donor	Gene expression reads		Surface protein reads	Cut-offs for quality control filtering (based on the Illumina sequencing data)				
Sample name	Illumina	Nanopore	Illumina	Features	Mitochondria (%)	CD45 reads	PE$^+$	PE$^-$
A	422,441,331	NA	73,536,058	$500 \leq$ Features ≤ 6500	≤ 8	$\geq 10^{2.5}$	$\geq 10^{2.9}$	$\leq 10^{2.2}$
B	392,640,438	NA	80,885,932	$500 \leq$ Features ≤ 3500	≤ 6	$\geq 10^{2.5}$	$\geq 10^{2.3}$	$\leq 10^{1.5}$
C	724,347,603	95,369,616	276,126,031	$500 \leq$ Features ≤ 3500	≤ 8	$\geq 10^{1.5}$	NA	NA
D rep1	786,665,522	71,790,574	260,983,738	$500 \leq$ Features ≤ 3500	≤ 8	$\geq 10^{1.5}$	NA	NA
D rep2	750,427,247	NA	351,590,581	$500 \leq$ Features ≤ 3500	≤ 8	$\geq 10^{1.5}$	NA	NA

Abbreviations: NA, not applicable; rep, replicate.

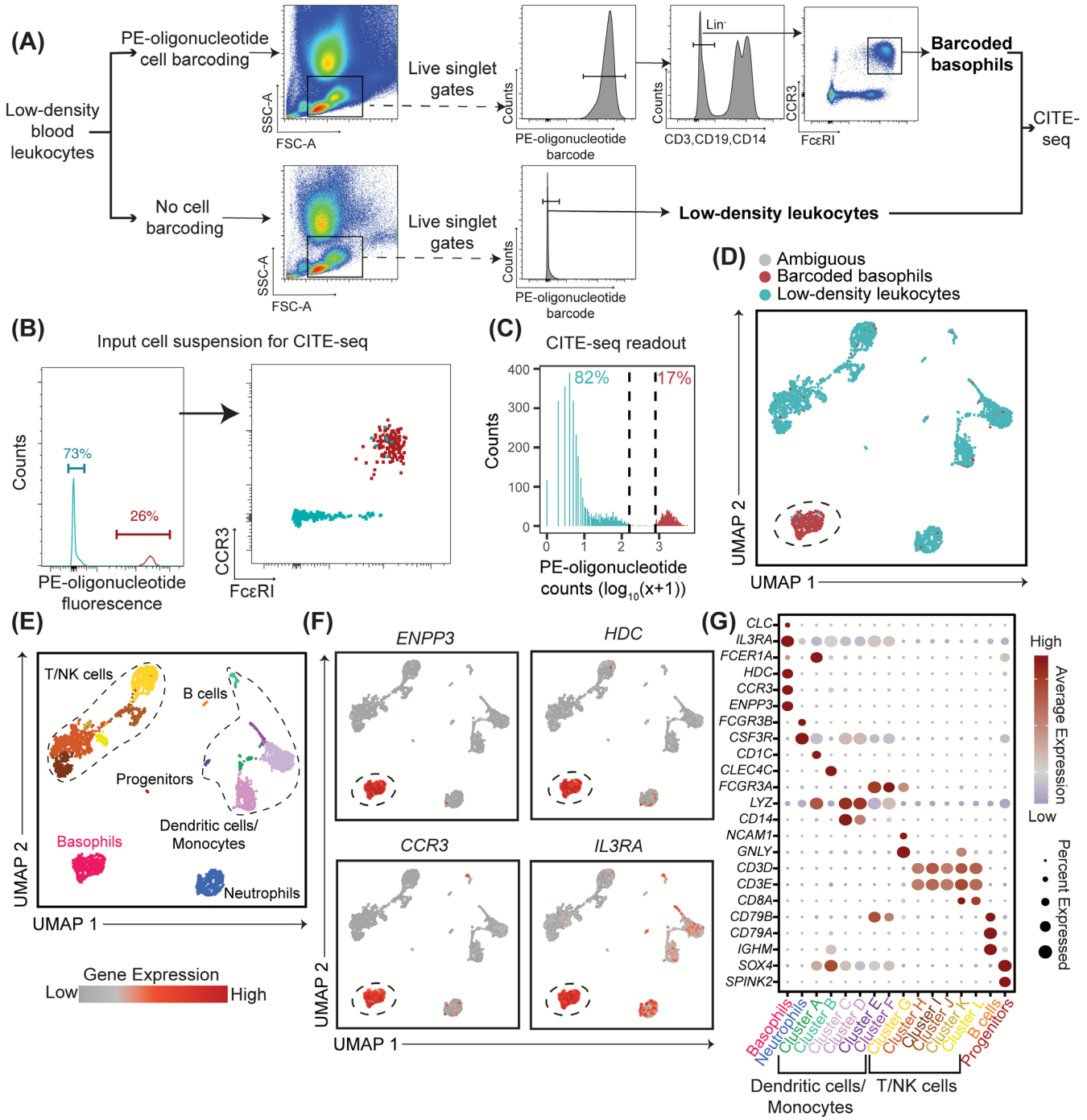

FIGURE 1 | Single-cell RNA-sequencing captures the gene expression profile of circulating basophils. (A) Experimental design. (B) Flow cytometry analysis showing the proportion of barcoded cells prior to CITE-seq analysis. The plots show live cells. (C) Analysis of CITE-seq data showing the proportion of barcoded cells. The histogram shows the distribution of the PE-oligonucleotide derived sequencing reads after the quality control and filtering steps. (D) UMAP visualization of the single-cell transcriptomics dataset (3720 cells), colored according to panel C. (E) Cells colored according to Leiden clusters at resolution 0.4. The clusters are annotated based on marker genes in panel G. (F) Normalized gene expression of selected genes. The basophil cluster is encircled. A comprehensive list of genes selectively expressed by basophils is shown in Table S1. (G) Dot plot generated with the DotPlot function of the Seurat package showing the scaled average expression of cell-type-related genes of the clusters shown in panel E. The colors of the x-axis labels correspond to the colors of the Leiden clusters shown in panel E.

identification of the basophil cluster (Figure 1D,E). Plotting the expression levels of cell-type-related genes annotated the remaining clusters, which included lymphocyte populations, dendritic cells/monocytes, and neutrophils (Figure 1E–G). As expected, the basophil cluster expressed genes coding for the basophil-related markers CD203c (*ENPP3*), histidine decarboxylase (*HDC*), CCR3 (*CCR3*), and the IL-3 receptor (*IL3RA*) (Figure 1F, Table S1) – consistent with basophils analyzed in bulk [11]. Taken together, our results revealed the successful analysis of basophils using CITE-seq.

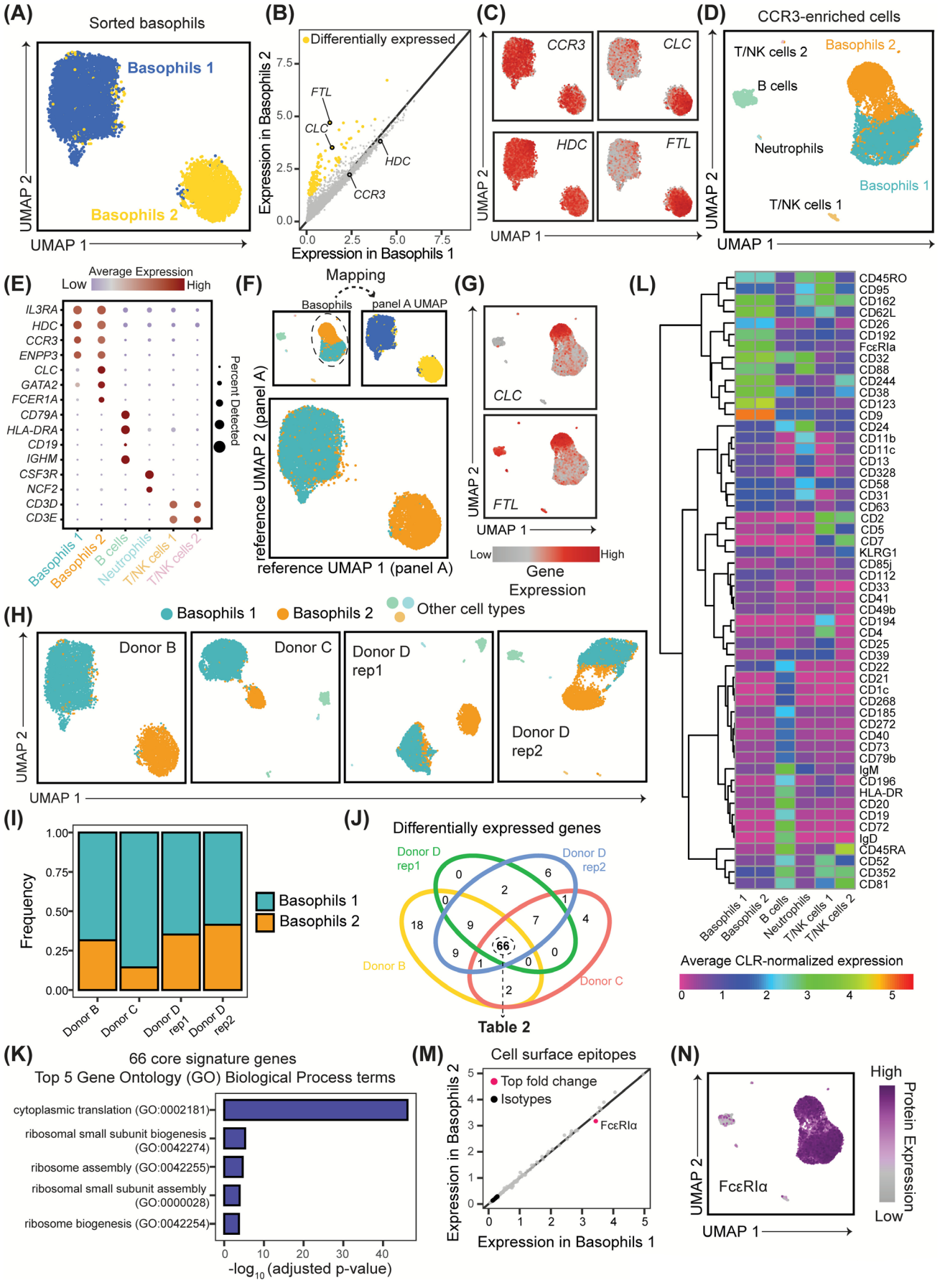

FIGURE 2 | Legend on next page.

FIGURE 2 | A multimodal resource reveals two transcriptionally distinct basophil populations with similar immunoprofiles. (A) UMAP visualization of 6285 sorted basophils colored by Leiden clusters (resolution 0.1). (B) Scatter plot showing the mean normalized gene expression of the two basophil populations on a $\log_2(x+1)$ scale according to the AverageExpression function. The colored dots represent differentially expressed genes. (C) UMAP visualizations showing the normalized gene expression (gray = low; red = high). (D) UMAP visualization of 15,795 CCR3-enriched cells from 2 donors and 3 CITE-seq libraries colored by annotated Leiden clusters (resolution 0.1). (E) Dot plot generated with the DotPlot function showing the scaled average gene expression of cell-type-defining genes used to annotate the cells in panel D. (F) Mapping basophils from panel D onto the UMAP space of the sorted basophils (panel A). (G) UMAP visualizations of the normalized gene expression. (H–J) Analysis of each dataset independently. (H) UMAP visualization colored by Leiden clusters (resolution 0.1). The leftmost plot is repeated from panel A for side-by-side comparison. (I) Population frequencies. (J) Venn diagram showing overlapping differentially expressed genes across datasets. All genes were upregulated in Basophils 2 versus Basophils 1. (K) Gene Ontology analysis. (L) Heatmap showing cell surface epitope expression based on the oligonucleotide-conjugated antibodies (hierarchical clustering). Markers shown are differentially expressed between basophils and other cell types. (M) Scatter plot showing mean CLR-normalized cell surface epitope expression of the two basophil populations. (N) UMAP visualization of the CLR-normalized protein expression of FcεRIα, the protein with the highest fold change.

3.2 | Single-Cell RNA-Sequencing Analysis Reveals Two Transcriptionally Distinct Basophil Populations

To investigate the transcriptional heterogeneity of basophils, we generated an additional dataset comprising basophils (side scatter^{low} lineage$^-$ CCR3$^+$ FcεRI$^+$ cells) as the only cell type (see methods and Figure S1D–G). Specifically, basophils were isolated using fluorescence-activated cell sorting, and the sorted cells were analyzed using CITE-seq. Strikingly, we identified two transcriptionally distinct clusters of basophils, referred to as Basophils 1 and 2 (Figure 2A). Further analysis revealed 105 differentially expressed genes that were higher in Basophils 2 compared with Basophils 1 (Figure 2B). Examples of upregulated genes included *CLC* and *FTL*, coding for galectin-10/Charcot-Leyden crystals and ferritin light chain, respectively. The expression levels of the basophil-associated genes *CCR3* and *HDC* were comparable between the clusters (Figure 2C).

The unexpected discovery of two transcriptionally distinct basophil clusters warranted further investigation and the generation of an additional dataset from independent donors. To assess the robustness of observing the two populations, we changed the cell isolation procedure from fluorescence-activated cell sorting-based to magnetic bead-based basophil purification. A CCR3-based isolation method was chosen to enrich for the basophil population, and density gradient centrifugation ensured that virtually no CCR3$^+$ eosinophils remained before the single-cell sequencing protocol (Figure S1H,I). The CCR3-enriched basophils were incubated with a panel of 140 different oligonucleotide-conjugated antibodies prior to the CITE-seq analysis to maximize the insights gained from the additional samples. A total of three replicates from two independent donors together formed a comprehensive multimodal dataset of basophils, with transcriptomics data coupled with large-scale immunoprofiling data of each single cell.

Transcriptome-based UMAP visualization revealed two basophil clusters in the integrated multimodal dataset (Figure 2D,E), which were not driven by individual replicates or cell cycle effects (Figure S1J). To investigate whether these clusters corresponded to the Basophils 1 and 2 clusters in the basophil-specific dataset (Figure 2A), we mapped the basophils from the integrated multimodal datasets onto this reference (Figure 2F). The clusters in the integrated multimodal dataset, also referred to as Basophils 1 and 2, mapped onto the respective clusters in the basophil-specific dataset (Figure 2F). This observation indicated that the two basophil clusters can be independently identified across datasets.

High expression of the *CLC* and *FTL* genes marked the Basophils 2 cluster across the basophil-specific dataset and the multimodal dataset (Figure 2C,G). To establish a comprehensive list of differentially expressed genes between Basophils 1 and 2, we performed a differential expression analysis—interrogating each donor and replicate separately to ensure robustness. Visualizing and clustering each donor and replicate revealed that Basophils 1 and 2 clusters were reproducibly distinguished across 4 replicates, 3 donors, and 2 cell preparation methods (Figure 2H, Figure S1K) – with Basophils 2 frequencies ranging between 14% and 41% (Figure 2I). In each dataset, 81–105 genes were differentially expressed between Basophils 1 and 2. Intersecting the lists from each dataset established a robust transcriptomic signature of 66 genes (Figure 2J, Table 2), all of which were higher in Basophils 2 compared with Basophils 1. Gene Ontology analysis identified biological processes related to translation and ribosome assembly and biogenesis (Figure 2K), reflecting that the list included a large collection of genes encoding ribosomal proteins. Notably, downregulation of genes encoding ribosomal proteins is associated with basophil maturation in mouse hematopoiesis [16]. We therefore speculate that Basophils 1 and 2 represent distinct maturation stages.

As alternative approach to analyze the single-cell transcriptomics data, we integrated the data from the 3 donors. UMAP visualization and clustering again identified the Basophils 1 and 2 populations across donors (Figure S2A–C). The sex of each anonymous donors was inferred based on the expression of *XIST* (female) and *DDX3Y* (male), confirming that both basophil populations were captured independent of donor sex (Figure S2D). Differential gene expression analysis between Basophils 1 and 2 identified 94 genes, which included all 66 genes from the previously defined signature (Figure S2E). In addition to the genes encoding ribosomal proteins upregulated in Basophils 2, the expanded list included *FCER1A* and *CXCR4*, which have been reported to be downregulated during basophil maturation in mouse [17]. These findings further support the interpretation that Basophils 2 represent a less mature population.

TABLE 2 | Differentially expressed genes (all higher in Basophils 2).

Genes					
ACTB	*FTL*	*PNRC1*	*RPL30*	*RPLP2*	*RPS6*
ATP5F1E	*GAPDH*	*PTMA*	*RPL32*	*RPS12*	*S100A6*
CEBPB	*H3F3B*	*RFLNB*	*RPL34*	*RPS14*	*SH3BGRL3*
CFD	*HLA-A*	*RGS2*	*RPL36*	*RPS20*	*TMSB10*
CLC	*IFITM2*	*RPL10*	*RPL37*	*RPS24*	*TMSB4X*
EEF1A1	*IRF2BP2*	*RPL11*	*RPL37A*	*RPS25*	*TOB1*
EIF1	*JUND*	*RPL12*	*RPL38*	*RPS26*	*TPT1*
ELOB	*MS4A3*	*RPL13*	*RPL39*	*RPS27*	*TXNIP*
EMP3	*MYL6*	*RPL13A*	*RPL41*	*RPS27A*	*UBA52*
FAU	*PCBP1*	*RPL21*	*RPL9*	*RPS28*	*UBC*
FTH1	*PET100*	*RPL28*	*RPLP1*	*RPS29*	*ZFP36*

Our CITE-seq-based approach characterized the expression of 134 cell surface markers and included 6 isotype controls. None of the 134 profiled markers distinguished Basophils 1 and 2 (Figure 2L–N, Figure S3). This included basophil activation markers, such as CD11b, CD44, CD54, CD63, and CD69 [18]. To further search for potential protein markers that separate basophils into the Basophil 1 and 2 states, we revisited the lists of differentially expressed genes (Figure S2E, Table 2) and performed flow cytometry analysis of 3 candidates: CXCR4, MS4A3, and galectin-10 (*CLC* gene). Basophils expressed all 3 markers at the protein level (Figure S4), in agreement with the gene expression data. However, none of the protein markers confidently captured basophil heterogeneity indicative of Basophils 1 and 2.

We next examined the cell surface marker expression independently of the transcriptome-based clusters to assess whether any marker could subset the basophils based on the surface phenotype alone. CD22, CD123, CD352, and FcεRI exhibited significant heterogeneity and bimodal expression patterns (Figure S5A). However, these bimodal expression patterns were largely driven by donor variability (Figure S5B). Notably, only CD352 subsets the basophils on the surface protein level within individual donors (Figure S5B). Nevertheless, the CD352 levels were comparable in the Basophils 1 and 2 clusters (Figure 2L, Figure S5C).

3.3 | Long-Read-Based CITE-Seq Analysis Profiles the Basophil Populations

Single-cell RNA-sequencing based on the Oxford Nanopore Technology (ONT) has the potential to capture full-length transcripts. We reasoned that this long-read sequencing technology could potentially improve the detection of specific genes, thereby gaining insights into basophil heterogeneity beyond the results obtained from the default short-read Illumina-based sequencing pipeline. Results based on the Nanopore platform would also enable us to validate the existence of basophil populations using an independent sequencing technology. To generate the long-read sequencing data, we produced Nanopore sequencing libraries from the existing PCR-amplified full-length cDNA libraries used for generating the results of the CCR3-enriched basophils presented in Figure 2D—one from each donor. The preservation of the cell barcodes on the cDNA sequences allowed comparative analysis between the Nanopore- and Illumina-based results (Figure 3A, Figure S6). UMAP visualization of the Nanopore data revealed three main basophil clusters, designated as Basophils 1, Basophils 1 ONT, and Basophils 2 (Figure 3B). Both Basophils 1 and Basophils 1 ONT mainly consisted of cells annotated as Basophils 1 in the Illumina-based analysis (Figure 3C). The discrimination of the Nanopore-specific Basophil 1 ONT cluster was attributed to two long non-coding RNAs (Figure 3D). The Nanopore Basophils 2 cluster mainly consisted of cells of the corresponding Illumina cluster (Figure 3B,C). Differential expression analysis based on the Nanopore data revealed higher expression of the characteristic *CLC* and *FTL* genes in Basophils 2 than in Basophils 1 (Figure 3E,F), in agreement with the Illumina-based results (Figure 2B,C,G, Table 2). We also detected higher levels of *TSPAN9* and *OOEP* in Basophils 2 than in Basophils 1 in the Nanopore dataset, genes that were poorly captured in the Illumina dataset despite its overall higher sequencing depth (Figure 3G,H, Table 1). These observations highlight the advantage of using two independent sequencing pipelines.

4 | Discussion

Here, we present a cellular and molecular blueprint of circulating human basophils, accessible through a user-friendly web interface (http://dahlinlab.cmm.se). The resource integrates RNA sequencing with large-scale immunoprofiling to enable data-driven exploration of basophil biology at the single-cell level. Using data from this resource, we identify and characterize two previously unrecognized basophil populations, laying the groundwork for future research on the basophil's role in health and disease.

Human basophils are challenging to capture using single-cell RNA-sequencing. To address this, the single-cell RNA-sequencing protocol, the sequencing depth, and the analysis

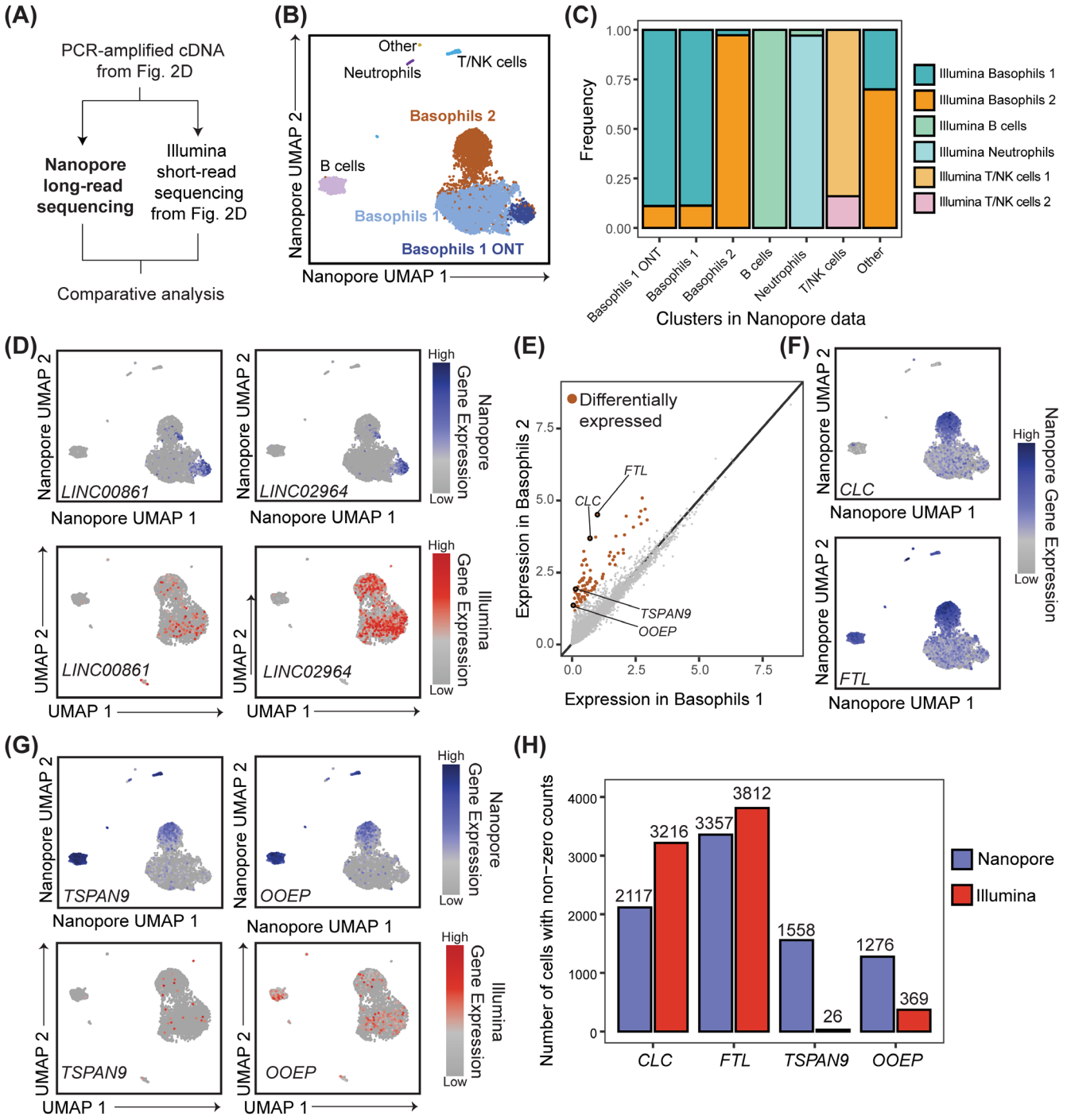

FIGURE 3 | Long-read single-cell RNA sequencing provides additional insights into the basophils' transcriptome. (A) Experimental design. (B) UMAP visualization of the long-read sequencing data showing 7304 cells from 2 donors. The colors indicate annotated Leiden clusters at resolution 0.1. (C) Bar plot showing the cell composition of the Nanopore-defined clusters (x-axis) according to the Illumina cluster annotation. (D) UMAP visualizations showing the normalized gene expression based on Nanopore and Illumina sequencing data. Note that only the cells that were profiled with both sequencing pipelines (7304 cells) are shown in the plots. (E) Scatter plot showing the Nanopore-based mean normalized gene expression in the Basophils 1 and 2 populations on $\log_2(x+1)$ scale according to the AverageExpression function. The colored dots highlight differentially expressed genes. (F–G) UMAP visualizations showing normalized gene expression based on Nanopore and Illumina sequencing data. Only the cells profiled by both technologies are shown. (H) Bar plot showing the number of cells with at least one sequencing read, plotted separately for each sequencing pipeline. Only the cells profiled with both technologies were interrogated.

pipeline were adapted accordingly (see Methods). Bioinformatics analysis identified two transcriptionally distinct basophil populations. Despite screening more than 100 different antibodies, no specific cell surface marker that allows the prospective isolation of each transcriptionally distinct population was found, making functional analysis difficult. However, the gene expression analysis revealed that Basophils 2 express higher levels of *FCER1A*, *CXCR4*, and genes encoding ribosomal proteins than Basophils

1. In mice, downregulation of such genes is associated with basophil maturation [16, 17]. We therefore speculate that Basophils 2 and Basophils 1 represent sequential maturation states, in which Basophils 1 are terminally differentiated. Nevertheless, this preliminary hypothesis requires future functional validation.

Basophil heterogeneity has previously been examined at the surface marker level using mass cytometry. Simultaneous measurement of 44 proteins identified 4 potential basophil states, distinguished based on the expression levels of CD16, CD244, and FcεRI [9]. Our single-cell omics-based resource included data on more than 100 cell surface markers, allowing for comprehensive analysis of the basophils' immunophenotype. In line with the mass cytometry results, we observed that basophils express heterogeneous levels of FcεRI. However, given that FcεRI expression on basophils correlates with serum IgE levels [19, 20], we further analyzed the bimodal expression pattern of FcεRI. Notably, this pattern was driven by donor variation, rather than reflecting stable basophil populations within individual donors. A single marker, CD352, displayed a bimodal expression pattern within individual donors. Flow cytometry analysis of CD352 verifies this observation [21]. However, the CD352 expression did not distinguish the transcriptionally distinct basophil populations and therefore reflects an independent type of heterogeneity.

We intentionally isolated basophils using two different approaches to assess the robustness of our findings on basophil heterogeneity. Notably, the identification of Basophils 1 and 2 populations and the main differentially expressed genes were consistent across isolation methods and donors. However, we cannot exclude that the isolation method has subtle effects on the population abundancies or gene expression profile. Furthermore, we cannot rule out the existence of additional basophil populations, for example, high-density basophils that may be lost during density centrifugation. Nevertheless, we robustly identified two distinct basophil populations, Basophils 1 and Basophils 2, across experimental approaches.

Peripheral blood basophils are clinically analyzed using the basophil activation test (BAT) to diagnose allergies. In this test, cell surface expression of CD63 following in vitro exposure to the relevant allergens is used as a proxy for basophil activation, indicating that the patient is sensitized. Notably, only a fraction of the basophils from a given patient typically becomes activated, resulting in a bimodal CD63 expression pattern that reflects activated and non-activated cell states [22]. Our data show that the CD63 cell surface expression is comparable between the Basophils 1 and 2 populations under steady-state conditions. However, whether the two populations differ in their capacity to become activated—contributing to the bimodal expression pattern—is an open question. Quantifying and determining the role of Basophils 1 and 2 in disease settings are also topics for future research. Thus, the presented single-cell resource serves as a foundation to investigate how basophil heterogeneity contributes to clinically relevant outcomes.

To conclude, single-cell short- and long-read-based RNA-sequencing coupled with large-scale immunoprofiling detects two transcriptionally distinct basophil populations. The single-cell omics data presented here constitutes a comprehensive reference resource—accessible through a user-friendly web interface (http://dahlinlab.cmm.se) – for future studies on basophils in health and disease.

Author Contributions

Sofia Papavasileiou and Joakim S. Dahlin conceptualized the study and designed the experiments. Sofia Papavasileiou, Jiezhen Mo, Lucille Margerie, and Chenyan Wu performed the experiments. Sofia Papavasileiou performed and led the data analysis. Daryl Boey and Magnus Tronstad contributed to the bioinformatics analysis and critically reviewed the analysis scripts. Daryl Boey established the web interface. Remi-André Olsen and Jörg A. Bachmann contributed to the long-read sequencing analysis. Jing Hui Low, Jocelyn Ong, Lars Heede Blom, and Anand Kumar Andiappan contributed to the bioinformatics analysis and data interpretation. Gunnar Nilsson contributed to critical discussions and data interpretation. Joakim S. Dahlin supervised the study and secured funding. Sofia Papavasileiou and Joakim S. Dahlin wrote the manuscript. All authors critically reviewed the manuscript.

Acknowledgements

The research was funded by grants from the Swedish Research Council (2022-00558), the Swedish Cancer Society, and Karolinska Institutet. Sofia Papavasileiou was supported by a scholarship from the Onassis Foundation (scholarship ID: F ZO 057-2/2022-2023). Anand Kumar Andiappan, Jing Hui Low, and Jocelyn Ong were supported by core funds from the Singapore Immunology Network, Agency for Science, Technology and Research (A*STAR). The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, and the Swedish Research Council, and SNIC/NAISS/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Sequencing data was processed using resources provided by NAISS, partially funded by the Swedish Research Council through grant agreement no. 2022-06725. ChatGPT was used to assist with language refinement during the preparation of the manuscript. All suggestions were carefully reviewed before implementation.

Funding

This work was supported by Vetenskapsrådet, Onassis Foundation, Singapore Immunology Network, Agency for Science, Technology and Research (A*STAR), Cancerfonden, and Karolinska Institutet.

Conflicts of Interest

Chenyan Wu is currently employed by Genoskin. Genoskin had no role in the study. All other authors declare no competing financial interests.

Data Availability Statement

Processed sequencing data are accessible through the Gene Expression Omnibus accession number GSE287976. Links to the online web resources are available through the accession number, and at http://basoscrna1.dahlinlab.cmm.se (Figure 1), http://basoscrna2.dahlinlab.cmm.se (Figure 2A), http://basoscrna3.dahlinlab.cmm.se (Figure 2D), and http://basoscrna4.dahlinlab.cmm.se (Figure 3B). Access to raw sequencing data is regulated by the Swedish Ethical Review Authority, Swedish legislation, and General Data Protection Regulation. Access to such data is provided by the corresponding author, given that all legal requirements are fulfilled.

References

1. J. E. Rood, S. Wynne, L. Robson, et al., "The Human Cell Atlas From a Cell Census to a Unified Foundation Model," *Nature* 637, no. 8048 (2025): 1065–1071, https://doi.org/10.1038/s41586-024-08338-4.

2. N. Ben-Baruch Morgenstern, M. Rochman, M. Kotliar, et al., "Single-Cell RNA-Sequencing of Human Eosinophils in Allergic Inflammation in the Esophagus," *Journal of Allergy and Clinical Immunology* 154, no. 4 (2024): 974–987, https://doi.org/10.1016/j.jaci.2024.05.029.

3. N. Iwasaki, J. A. Poposki, A. Oka, et al., "Single Cell RNA Sequencing of Human Eosinophils From Nasal Polyps Reveals Eosinophil Heterogeneity in Chronic Rhinosinusitis Tissue," *Journal of Allergy and Clinical Immunology* 154, no. 4 (2024): 952–964, https://doi.org/10.1016/j.jaci.2024.05.014.

4. J. M. Rodrigo-Munoz, S. Naharro-Gonzalez, S. Callejas, et al., "Single-Cell RNA Sequencing of Human Blood Eosinophils Reveals Plasticity and Absence of Canonical Cell Subsets," *Allergy* 80, no. 2 (2025): 570–574, https://doi.org/10.1111/all.16213.

5. E. Ronnberg, A. Ravindran, L. Mazzurana, et al., "Analysis of Human Lung Mast Cells by Single Cell RNA Sequencing," *Frontiers in Immunology* 14 (2023): 1151754, https://doi.org/10.3389/fimmu.2023.1151754.

6. N. Ben-Baruch Morgenstern, A. Y. Ballaban, T. Wen, et al., "Single-Cell RNA Sequencing of Mast Cells in Eosinophilic Esophagitis Reveals Heterogeneity, Local Proliferation, and Activation That Persists in Remission," *Journal of Allergy and Clinical Immunology* 149, no. 6 (2022): 2062–2077, https://doi.org/10.1016/j.jaci.2022.02.025.

7. D. F. Dwyer, J. Ordovas-Montanes, S. J. Allon, et al., "Human Airway Mast Cells Proliferate and Acquire Distinct Inflammation-Driven Phenotypes During Type 2 Inflammation," *Science Immunology* 6, no. 56 (2021): eabb7221, https://doi.org/10.1126/sciimmunol.abb7221.

8. E. J. Leonard and A. Skeel, "Separation of Human Basophils Into Two Fractions With Different Density and Histamine Content," *Journal of Allergy and Clinical Immunology* 76, no. 4 (1985): 556–562, https://doi.org/10.1016/0091-6749(85)90775-4.

9. N. Vivanco Gonzalez, J. P. Oliveria, D. Tebaykin, et al., "Mass Cytometry Phenotyping of Human Granulocytes Reveals Novel Basophil Functional Heterogeneity," *iScience* 23, no. 11 (2020): 101724, https://doi.org/10.1016/j.isci.2020.101724.

10. M. Stoeckius, C. Hafemeister, W. Stephenson, et al., "Simultaneous Epitope and Transcriptome Measurement in Single Cells," *Nature Methods* 14, no. 9 (2017): 865–868, https://doi.org/10.1038/nmeth.4380.

11. G. Monaco, B. Lee, W. Xu, et al., "RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types," *Cell Reports* 26, no. 6 (2019): 1627–1640.e7, https://doi.org/10.1016/j.celrep.2019.01.041.

12. Y. Hao, S. Hao, E. Andersen-Nissen, et al., "Integrated Analysis of Multimodal Single-Cell Data," *Cell* 184, no. 13 (2021): 3573–3587.e29, https://doi.org/10.1016/j.cell.2021.04.048.

13. D. MacGlashan, Jr., "Development of a Microarray-Based Method to Detect Exposure of Human Basophils to IL-3," *Journal of Immunological Methods* 385, no. 1–2 (2012): 51–59, https://doi.org/10.1016/j.jim.2012.08.006.

14. D. MacGlashan, "Expression Profiling of Human Basophils: Modulation by Cytokines and Secretagogues," *PLoS One* 10, no. 5 (2015): e0126435, https://doi.org/10.1371/journal.pone.0126435.

15. S. U. Sonder, M. Plassmeyer, D. Loizou, and O. Alpan, "Towards Standardizing Basophil Identification by Flow Cytometry," *Frontiers in Allergy* 4 (2023): 1133378, https://doi.org/10.3389/falgy.2023.1133378.

16. F. K. Hamey, W. W. Y. Lau, I. Kucinski, et al., "Single-Cell Molecular Profiling Provides a High-Resolution Map of Basophil and Mast Cell Development," *Allergy* 76, no. 6 (2021): 1731–1742, https://doi.org/10.1111/all.14633.

17. K. Miyake, J. Ito, J. Nakabayashi, S. Shichino, K. Ishiwata, and H. Karasuyama, "Single Cell Transcriptomics Clarifies the Basophil Differentiation Trajectory and Identifies Pre-Basophils Upstream of Mature Basophils," *Nature Communications* 14, no. 1 (2023): 2694, https://doi.org/10.1038/s41467-023-38356-1.

18. C. Yoshimura, M. Yamaguchi, M. Iikura, et al., "Activation Markers of Human Basophils: CD69 Expression Is Strongly and Preferentially Induced by IL-3," *Journal of Allergy and Clinical Immunology* 109, no. 5 (2002): 817–823, https://doi.org/10.1067/mai.2002.123532.

19. F. J. Malveaux, M. C. Conroy, N. F. Adkinson, Jr., and L. M. Lichtenstein, "IgE Receptors on Human Basophils. Relationship to Serum IgE Concentration," *Journal of Clinical Investigation* 62, no. 1 (1978): 176–181, https://doi.org/10.1172/JCI109103.

20. S. S. Saini, A. D. Klion, S. M. Holland, R. G. Hamilton, B. S. Bochner, and D. W. Macglashan, "The Relationship Between Serum IgE and Surface Levels of FcepsilonR on Human Leukocytes in Various Diseases: Correlation of Expression With FcepsilonRI on Basophils but Not on Monocytes or Eosinophils," *Journal of Allergy and Clinical Immunology* 106, no. 3 (2000): 514–520, https://doi.org/10.1067/mai.2000.108431.

21. L. H. Blom, E. A. Bartko, T. K. R. Moller, L. K. Poulsen, and B. M. Jensen, "FcepsilonRI-Activated Basophils Express CCR4, CCR8, CCR9, CCX-CKR and XCR1," *Allergy* 78, no. 2 (2023): 539–543, https://doi.org/10.1111/all.15488.

22. D. G. Ebo, C. H. Bridts, M. M. Hagendorens, N. E. Aerts, L. S. De Clerck, and W. J. Stevens, "Basophil Activation Test by Flow Cytometry: Present and Future Applications in Allergology," *Cytometry Part B: Clinical Cytometry* 74, no. 4 (2008): 201–210, https://doi.org/10.1002/cyto.b.20419.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supplementary Methods Figure S1:** Detection of basophils among low-density leukocytes. **Figure S2:** Integrated analysis combining data from three donors. **Figure S3:** The two basophil populations have similar immunophenotypes. **Figure S4:** Flow cytometry analysis of leukocytes. **Figure S5:** Immunoprofiling of the circulating human basophils. **Figure S6:** Identification of transcriptional heterogeneity by long-read sequencing. **Table S1:** all70209-sup-0002-TableS1.csv.